ETS *Educational Testing Service*

K-12 Assessments

Guidelines

for

SCIE
ENGLISH
MATHEMATICS
HISTORY

Item

Writers

K-12 Assessments
1250 NE Loop 410 Suite 1000, San Antonio, TX 78209

Table of Contents

Introduction

This booklet has been prepared by the staff of Educational Testing Service to give item writers guidelines for writing test items for students in kindergarten through grade 12.

Writing test items is an important endeavor, since K-12 test scores usually have very high stakes for all participants in the educational process—students, educators, parents, and policymakers. As an item writer for standards-based assessments, it is vital that you possess a strong knowledge of the content area to be tested. It is also important you know the principles of good item writing. As you write, you should keep in mind the guidelines in this booklet. You should also get as much feedback as possible on your items. Item writing expertise, like all writing expertise, is developed through conscientious practice of the elements of good writing.

Here are some general considerations for you to keep in mind as you write:

Do original work. Ideas for items will come from your teaching experience and content knowledge, but all the test questions you write must be your original work. For example, borrowing a math context from a textbook and writing a new item for it is plagiarism. Selecting a passage from a publisher's sample test materials is also unacceptable.

Be creative. Although items must fit into a framework of specific educational objectives as well as the guidelines in this booklet, items should be as interesting as possible to students.

Remember your audience—the students who will read your items—at all times. Your objective should be to give students an opportunity to demonstrate what they have learned and can accomplish in a particular content area.

Enjoy the opportunity. We hope that you will find that writing test items is a challenging and exciting professional activity.

1

About ETS

Educational Testing Service is the world's largest private educational testing and measurement organization and a leader in educational research. A nonprofit company dedicated to serving the needs of individuals, educational institutions and agencies, and governmental bodies in 181 countries, ETS develops and annually administers more than 12 million tests worldwide on behalf of clients in education, government, and business.

ETS was founded in 1947 after it became clear that an increased demand for the SAT® test, GRE® test, and various achievement tests required the creation of an operating staff considerably larger than the organizations responsible for the tests at the time. On December 19, 1947, after a series of negotiations, the American Council on Education, the Carnegie Foundation for the Advancement of Teaching, and the College Entrance Examination Board turned over their testing programs, a portion of their assets, and some of their employees to form ETS—a single national organization devoted exclusively to educational testing and research.

The Educational Measurement and Test Development department of K-12 Assessments works with state education agencies, teachers, and school administrators to create tests that measure student learning in kindergarten through grade twelve. These tests are custom designed to measure what students should know and be able to do, according to the state curriculum or educational standards. This department collaborates with customers at every step in the development process, from conception through final publication and implementation, to create tests that are valid, reliable, fair, and that mirror proven instructional practices.

The departmental goal is to create tests that provide valuable information for students, parents, teachers, and policy makers that ultimately help improve the quality of education available to all students. The professional staff members of Educational Measurement and Test Development have many years of experience as educators and test developers. They are motivated to develop tests consistently of the highest quality and, at the same time, to provide tests that are designed to be positive learning experiences for all students.

The Development Process

While each testing program is unique, there are some common steps in the development process.

1. Decide how the test should be designed and what should be assessed

In this step of the development process, the Test and Item Specifications are created. The Test Specifications give information on the test as a whole. Typically, it describes the purpose of test, how many items and what kinds will be on the test, and how the test will be scored. The Item Specifications give information about what is to be tested and how the items are to be developed. The standards that will be measured are identified, described, and grouped into domains. The document goes into great detail for each standard on what the items should measure and provides guidelines for developing the items. It is critical that item writers read and understand this document. All items for the testing program must follow the guidelines set forth in the Item Specifications.

2. Create Items

After the Test and Item Specifications have been developed, item development can begin. It is critical that the items adhere to the specifications for the testing program. Item writers are identified and trained for the testing program. Once the writers understand what is required, they are given their writing assignments.

3. Internally Review Items

Once an item has been accepted, it goes through several reviews. At each review, the item is reviewed for its match to the standard, the content accuracy and relevance, the grade level appropriateness, for any potential bias and sensitivity concerns, and for editorial considerations. The items must be well-written and adhere to the guidelines in the Item Specifications. The checklists in this guide help identify many of the things considered when an item is reviewed.

4. Externally Review Items

Once the item has been revised according to the comments from the internal reviews, the item is submitted to outside teacher review groups that review the item for content concerns and potential bias and sensitivity concerns. The item is again revised according to the comments from the external review.

5. Field Test Items

Items that have made it completely through all the review cycles are placed on a field test. The field test is administered to students. The students do not receive scores because the field test is testing the test NOT the students. The purpose of the field test is to get data on each item to help determine which items worked well and could be eligible to be used on an operational test.

6. Review Item Data

Each item on the field test is statistically analyzed to determine how well the item performed. Every aspect of the item is scrutinized.

7. Select Items

Based on the data review, items that did not perform well are removed from the pool of potential items for the operational test. From the items that remain, items are selected for use on the operational test.

8. Create the Operational Test

The operational test is built according to the Test Specifications using items from the item pool. The operational tests are then used in the testing program to assess students.

Writing Multiple-Choice Items

Guidelines for the Item as a Whole

- An item should measure the objective it is designed to measure (e.g., if the objective is to understand historical change, the item must address some kind of change, rather than the identification of a historical event or understanding of a historical fact). Label each item with the standard and specific objective to which you are writing.

- An item should have one and only one clearly correct answer.

- An item should be written to focus on one concept or skill that is clearly presented in the stem.

- Keep the cognitive level of the item clearly in mind. If the objective calls for higher order thinking, be sure that the item does not require simple recall to be answered correctly. Conversely, if the objective calls for the student to demonstrate factual knowledge, the item should have the appropriate cognitive level.

- Items should be at the appropriate level of difficulty for the grade level or course. The difficulty of an item should depend largely on the fineness of the distinctions that must be made to choose the correct answer, not on the obscurity of the information.

- Stems need to be clear and concise, and nonfunctional words must be avoided.

- While items need to be engaging, they also need to be written at the appropriate vocabulary level. Items should make use of vocabulary and sentence structures appropriate to the grade level of the students being tested. The reading level of items should be appropriate to the grade level or course, so that a student's knowledge and skills are not confounded with his or her reading ability.

- Avoid using vocabulary and idiomatic phrases that could be unfamiliar to students, especially those for whom English is a second language.

- Item writers must carefully avoid creating questions that draw upon students' outside knowledge, unless the purpose of an item is to test outside knowledge (e.g., in science or social studies). Keep in mind that students within one state can have very different backgrounds of experience.

6

- Items should not focus on trivialities. The knowledge or facts being tested should be worth testing.

- Items should reflect current, sound instructional practices.

- Items should never attempt to trick students, nor should they contain cute or clever distracters.

- Items should not appear to ask for a student's opinion (e.g., **Why do you think . . .**). Any answer could be defended as correct. Similarly, items that call for the student to draw a conclusion should specify the source the student should use for drawing the conclusion (e.g., **According to the article, why are forest fires sometimes beneficial?**).

- Watch for grammatical correctness in test items. For example, do not use anyone or everyone as an antecedent for they (e.g., **Why did everyone in the orchestra expect that they would win the contest?**). Many times an alternate construction can replace the indefinite pronoun (e.g., **Why did the members of the orchestra expect they would win the contest?** or **Why did everyone in the orchestra expect to win the contest?**). Remember that compound subjects carry plural verbs. For example, **The earl's wisdom and knowledge was important to the king because**— is incorrect. Also avoid dangling participles, run-on sentences, and fragments.

- Use the active voice in items, and avoid informal diction and usage unless quoting from an informal passage or other stimulus (e.g., avoid contractions unless they are being tested; do not split infinitives).

- Any factual information in the item must be accurate. Source material must be provided.

- In general, the *Chicago Manual of Style* should be consulted for style questions. For the specific style for capitalization and punctuation of the stem and response options, consult the test item specifications and style guide for the program for which you are writing.

- It is crucial that items be free from problems of bias or sensitivity. See the section of this document titled "Bias and Sensitivity Issues."

Guidelines for Multiple-Choice Stems and Stimuli

- Item stems should contain a complete statement of the idea being tested. For example, **An atom is—** is a poor stem because the student cannot know what kind of knowledge about an atom will be required: definition? function? purpose? A better stem might be **How is an atom different from an ion?**

- Whenever possible, the stem should contain a question, but incomplete statements (e.g., **Maria left the room because she decided to—**) are acceptable as well. "Fill in the blank" stems are not to be used (e.g., **The perimeter of the driveway is _____ meters**), except in certain kinds of English language items.

- Negative stems (e.g., use of **not,** except) should be avoided because these items can be especially difficult for students. If a negative is used in the stem, there should be no negatives in the distracters.

- Use the words **what** and **which** with care. What is usually used when there is only one possible existing answer (e.g., **What did President Roosevelt call his program for economic recovery after the Great Depression?**). Which usually implies that there are several possible ways to express the answer (e.g., **Which statement best describes the Roaring Twenties?**).

- With respect to graphs, charts, or other graphic materials, consider the following:

  - Is the stimulus necessary to answer the question, or is it merely decorative or supplementary?

  - Is the stimulus likely to be interesting to students?

  - Is the stimulus labeled correctly?

  - Is all of the information needed to answer the question(s) provided by the stimulus?

  - Is the stimulus clear and simple enough that it can be reproduced in a double-column test book or on a computer screen?

8

Guidelines for Response Options

- No response option (correct answer or distracter) should be significantly different from the other three options (e.g., in length, specificity, complexity, compound structure). However, pairs of options may reflect minor differences (e.g., two long and two short options).

- The response options should relate to the stem in the same way both grammatically and conceptually.

- No response option (correct answer or distracter) should eliminate another response option. Thus two distracters should not be opposites of each other. Savvy test takers will immediately narrow the choices to the two opposites and choose one of them.

- No response option (correct answer or distracter) should deny the truth of the stem (e.g., the stem asks **Why did Sophia go to the library?** and one option is **She did not go to the library.**).

- Response options should not give clues to students. Avoid the use of absolutes (e.g., never, always, all). Words used in the stem should not be repeated in the response options (correct answer or distracters).

- Avoid repeating words or phrases in the options that could be placed in the stem.

- The distracters should be plausible and reasonable in terms of the stimulus, the item stem, and the student's level of knowledge.

- Distracters may include a) common misconceptions or misunderstandings, or b) misreading of the stimulus. For some tests (particularly math), it is helpful for item writers to include the rationales (reasons) for the distracters.

- Do not use distracters that have nearly the same meaning even though they are worded differently (e.g., **Gwen was happy to go. Gwen was looking forward to the trip.**).

- In a set of items, make sure that there is a nearly equal number of A, B, C, and D choices.

Guidelines for Sets of Items

- Items based on a stimulus or passage should be independent of each other. The correct answer to one item should not depend upon a previously answered item.

- The longer or more complex the passage or stimulus, the greater number of items should be associated with it. It is not fair for students to have to spend time understanding a long or complex stimulus that is accompanied by only a few items.

- It is important that the items written for a single objective or group of objectives represent a range of difficulty. Keep in mind that you are writing items for the full spectrum of ability. Each set of items should contain a majority of items of medium difficulty and about 20 percent of the items at a more difficult level and about 20 percent at an easy level.

Common Errors in Writing Multiple-Choice Items

The following are examples of some of the most common item writing errors for multiple-choice test questions.

1. **Item stems that are too lengthy or imprecise.**

   Example:
   Poor: Which one of the following is something that can be found in a short story that is also found in this personal narrative?

   Corrected: Which of these short story elements is present in this personal narrative?

   Example:
   Poor: During one week, Callie spent $6\frac{3}{4}$ hours working on a project for her science class and $4\frac{1}{2}$ hours working on a book report on her favorite mystery book. She wanted to find out how much more time she spent on science than she spent on the book report. How many more hours did she spend on the science project?

   Corrected: Callie spent $6\frac{3}{4}$ hours working on a science project and $4\frac{1}{2}$ hours writing a book report. How much more time did she spend on the science project than on the book report?

2. **Item stems that are choppy or abstract.** Test takers need to have a clear picture of their task *before* they begin reading the answer choices.

   Example:
   Poor:
   Reading this chart —
   A*   can tell someone the growth of four stocks over a four-year period

   Corrected:
   This chart shows the growth of —
   A*   four stocks over a four-year period

11

3. **Item options that lack balance.** This problem can be evidenced in several ways, including "stacked" options, options that are dramatically different in "flavor," and one option that is conspicuous in some way.

Example:
Poor: Animal tracks can best be found in —
A*   mud or snow   ("Stacked" option—it contains two elements.)
B   grass
C   water
D   the library   (Option is dramatically different in "flavor.")

Option A is "stacked"—it tends to attract test takers because it contains two elements.
Option D is too different in "flavor"—test takers will quickly dismiss the option as not remotely viable. All options should have credibility and attractiveness.

Corrected: Animal tracks can best be found in —
A*   mud
B   grass
C   water
D   sand

4. **Item stems and options that needlessly repeat words.**

Example:
Poor: Penguins cannot fly because —
A*   their wings are too small
B   their wings are too flat
C   their wings are too curved
D   their wings are too heavy

Corrected: Penguins cannot fly because their wings are too —
A*   small
B   flat
C   curved
D   heavy

12

5. **A stimulus, item stem, or an option containing information that clues the correct answer of the item itself or of another item in the test.**

   Example:
   Stimulus: Ranger didn't like playing with the big neighborhood dogs. They ran fast and often left Ranger behind. One day Ranger found some play companions that <u>werent</u> any larger than he was. Now he is the leader of his pack.

   How should <u>werent</u> be written?
   A*   weren't
   B   werent'
   C   were'nt
   D   wer'ent

   The second word of the stimulus, <u>didn't</u>, clues this item because savvy test takers can match up the placement of the apostrophe in <u>didn't</u> with the apostrophe needed in <u>weren't</u>. One correction would be to write out the contraction <u>didn't</u> as two words. Another option would be to test this skill in a different stimulus where clueing would not result.

6. **Items that measure more than one standard.**

   Example:
   Stimulus: Holly sitted next to the band leader.

   How should this sentence be written?
   A*   Holly sat next to the band leader.
   B   Holly sit next to the band leader.
   C   Holly sitting next to the band leader.
   D   holly sat nexst to the band leader.

   In this case, the item was intended to measure the standard "Use correct verb formation." In this item, options A, B, and C provide various forms of the verb. Option D, however, tests capitalization and spelling (which are measured under separate standards). Therefore, this item is assessing a student's ability to do more than merely choose the correct verb in the sentence. For this reason, option D would have to be revised (so that only the verb contains the error) in order for this item to match the standard.

CHECKLIST FOR MULTIPLE-CHOICE ITEMS

✓	REQUIREMENTS
The item as a whole—	
	Measures the objective.
	Has one and only one clearly correct answer.
	Clearly presents one central idea.
	Has a clear purpose.
	Is at the correct level of difficulty.
	Contains simple, direct, unambiguous language.
	Uses age-appropriate vocabulary and sentence structures.
	Does not use vocabulary and idiomatic phrases that could be unfamiliar.
	Does not rely on students' possessing outside knowledge.
	Tests worthwhile (not trivial) concepts or information.
	Follows the test specifications.
	Reflects good and current teaching practices.
	Is not tricky or cute.
	Does not appear to ask for the student's opinion.
	Is grammatically correct.
	Uses the active voice and avoids informal diction and usage.
	Is factually accurate.
	Is free of bias or sensitivity problems.
The stem for the item—	
	Gives the test taker a full sense of what the item is asking.
	Is either a question or an incomplete statement.
	Is both clear and concise.
	If it contains a negative, contains no negatives in the distracters.
	Uses the words *what* and *which* accurately.

14

✓	**REQUIREMENTS**
The stimulus for the item—	
	Is required in order to answer the associated item(s).
	Is likely to be interesting to students.
	Is correctly and clearly labeled.
	Provides all the information needed to answer the items.
	Can be reproduced clearly in a test book.
The response options—	
	Are written so that no one option is significantly different from the others.
	Relate to the stem in the same way.
	Do not contain an option that denies the truth of any other option.
	Do not deny the truth of the stem.
	Do not give clues to students, such as the use of absolutes.
	Do not repeat words that could be placed in the stem.
	Include plausible and reasonable misconceptions and errors.
	Do not include distracters that are phrased differently but have the same meaning.
	Have a balance of A, B, C, and D responses.
The set of items—	
	Is numerous enough to justify the time required to read the stimulus.
	Contains items that are entirely independent of each other.
	Reflects a range of difficulty appropriate to the grade level.

Writing Constructed-Response Items

Developing an effective constructed-response (CR) item requires a consideration of the scoring rubrics that will be applied to students' responses. Here are some guidelines for developing effective CR items.

- Be sure to consult the test specifications for the program. The test specifications may include the generic rubric to be used as well as some sample items.

- Remember that constructed-response items should be used to measure objectives that could not be assessed easily in a multiple-choice item.

- The CR items should be designed to challenge students to think rather than to provide a rote response.

- Write the question or task as clearly as possible. Students should be able to understand exactly what the question or task means.

- The CR item should invite and support a range of responses, including a full and complete response that can be written in a short amount of time, a moderately complete response, and a minimal response. To ensure that the CR item will provide a range, write out a response for at least three levels.

- Be specific about the expected response (e.g., **Give at least two reasons why Jason decided to walk home.** rather than **Why did Jason make his decision?**). Directions to students should be consistent with the criteria used in the scoring rubric (e.g., if a rubric emphasizes description, then the item should direct students to provide descriptive details).

- Do not ask questions that appear to advocate a particular cultural concept or value that may not be appropriate for all students taking the test.

- Avoid questions that will invite personal responses about students' lifestyles, values, or beliefs. CR items may ask for a student's opinion with regard to the content of a stimulus, but the question must be able to be answered with information from the stimulus.

- Use age-appropriate vocabulary and sentence structures.

- Avoid using vocabulary and idiomatic phrases that could be unfamiliar to students, especially those for whom English is a second language.

- It is crucial that items be free from problems of bias or sensitivity. See the section of this document titled "Bias and Sensitivity Issues."

16

Common Errors in Writing Constructed-Response Items

The following are examples of some of the most common item writing errors for constructed-response test questions.

1. **Items that are intrusive.** Constructed-response items should allow the student to demonstrate achievement of the appropriate standard (e.g., writing skills, mathematical reasoning, response to literature) without inviting a discussion of personal beliefs and practices. The question should not ask the student to reveal personal information.

   Example:
   Poor: In this story, Juliet uncovers a family secret. Imagine that you had this experience. Compare the way Juliet reacted to the secret with the way you would have reacted. Use details from the story to support your answer.

   Corrected: In this story, Juliet uncovers a family secret. What does her reaction to this secret show the reader about her personality? Use details from the story to support your answer.

2. **Items that are too broad in scope.** Such items would be likely to elicit responses that are too general or are unrelated to the story. When a constructed-response item has an appropriate scope, students who understand the question are able to write a full and correct response within a reasonable amount of time (e.g., 15 minutes).

   Example:
   Poor: One of the themes in this story is "Nature should be experienced by everyone." Why is it important to spend time outdoors? Use details from the story to support your answer.

   Corrected: One of the themes in this story is "Nature should be experienced by everyone." In what ways is the plot affected by having the events take place outdoors? Use details from the story to support your answer.

17

3. **Items that are too narrow in scope.** Such items would be likely to elicit responses that are very specific and short. These items usually have a single answer (a "right" answer) and sometimes could be more effectively written in multiple-choice format. The responses to items that are too narrow will tend to fall into two categories: right (highest score point) or wrong (lowest score point). Usually it is desirable for constructed-response questions to elicit a range of responses so that students with partial understanding or skills receive partial credit.

Example:

Poor: What is the setting of this story? *(Anticipated Response: Summer in Miami.)*

Corrected: What are two ways the author's choice of setting in this story affect the characterization? *(The response could identify the setting and speak to such issues as motivation, character interaction, or use of dialect.)*

18

CHECKLIST FOR CONSTRUCTED-RESPONSE ITEMS

✔	REQUIREMENTS
The item—	
	Directly assesses the skills, knowledge, or abilities specified for the test.
	Clearly tells students what they are being asked to do.
	Uses precise action verbs and descriptive words.
	Invites and supports a range of responses.
	Is specific about the expected level of detail required in the response.
	Does not advocate a particular value that may not be common to all students.
	Does not invite personal responses about students' lives or values.
	Uses age-appropriate vocabulary and sentence structures.
	Does not use vocabulary and idiomatic phrases that may not be familiar.
	Is free from problems of bias or sensitivity.

Specific Guidelines for Mathematics Items

- Most mathematics items are written in context, and this context may be fiction (e.g., **How many students in the class like chocolate ice cream best?**) or nonfiction (e.g., **How much higher is Mount Everest than Mount McKinley?**). Whether fiction or nonfiction, contexts should be inherently logical and realistic. For example, a rectangular kitchen table is unlikely to have an area of 50 square feet.

- There should be a logical reason for doing the mathematics in the item (e.g., it may be logical to convert inches of rain into feet of snow, but it is not logical to describe rainfall in feet and convert it to inches).

- The reading level for mathematics items should be at or below the grade level of the students taking the test. Stems should be as concise as possible. Long stems may be broken up with diagrams or graphs. Students should be asked to demonstrate their knowledge of mathematics, rather than their reading ability, when answering mathematics items.

- Use precise mathematical terms (e.g., **line segment** rather than **line, dollars per hour** rather than **dollars an hour**).

- When an item is not testing computation, the numbers should be easy to compute. For example, if an item is testing a sixth grade student's understanding of the formula for volume, the dimensions should be given in whole numbers.

- Numerals used in the calculations must be represented numerically, not spelled out (e.g., **There were 4 balloons in Mrs. Sosa's hand. Only 1 was pink.**).

- Items should recognize that the representation of figures is inexact (e.g., **Which of these most closely resembles (or appears to be) a right angle?**).

- For multiple-choice items, numerical response options should be placed in ascending or descending order, unless doing so would tend to clue the correct answer.

- In multiple-choice items, one response option should not be conspicuously larger or smaller than the other three options.

- All options in a multiple-choice item should be plausible.

- If the response options in multiple-choice items are numerical, item writers should provide a clear rationale for each option. If a short-answer item has a numerical answer, item writers should supply the computation required to reach the answer.

20

Specific Guidelines for Science Items

- For multiple-choice items, all responses should represent plausible responses within the scientific context (e.g., no obvious impossibilities).

- If appropriate, response options for multiple-choice items should be placed in a logical order.

- Any data or other stimuli presented as true should be verified in two independent sources.

- Multiple-choice items should be difficult because they required a student to make fine distinctions to demonstrate mastery of the concept. They should not be difficult because the content is obscure or trivial.

Specific Guidelines for Social Studies Items ═══

- For multiple-choice items, all responses should represent plausible responses within the social studies context (e.g., no obvious historical impossibilities).

- If appropriate, response options for multiple-choice items should be placed in a logical order.

- Any data or other stimuli presented as true should be verified in two independent sources.

- Multiple-choice items should be difficult because they required a student to make fine distinctions to demonstrate mastery of the concept. They should not be difficult because the content is obscure or trivial.

Specific Guidelines for English Language Arts Items

- Students must read the passage in order to answer the questions.

- For passage-related items, the distracters should reflect information in the passage. The student should choose an incorrect option because it represents a misreading of the stimulus. Generally, information from a passage is paraphrased, rather than repeated verbatim.

- For passage-related items, the order of the response options must follow the order in which they occur in the passage.

- Many programs prefer that item writers avoid "which of the following" and use "which of these" instead.

- For items that ask students to choose relevant details, be certain that the response options for these items contain details rather than generalizations.

- For language items, be careful that the correct option in one item does not clue the answer in another (e.g., do not test capitalization in Sunday if in another part of the passage the word Monday is used).

- For capitalization items, use only one capitalization error in the stem (e.g., **The Geographical Studies institute sponsored a visit to the Grand Canyon.** In other words, use the choices Geographical studies institute and geographical studies institute as distracters, not in the stem).

- For items asking students to identify appropriate resources, vary the questions so that encyclopedia is not always the correct answer.

- For items regarding tables of contents, use chapters starting with even numbers.

- If the response options are page numbers, list them in numerical order.

- For items assessing planning strategies, use graphic organizers but be certain that only one correct answer exists.

Thinking Skills

A test should measure a range of thinking skills. There are several models for ranking thinking skills. Testing programs will often specify which classification should be used and require that each item be labeled according to the thinking skill required to answer it.

One popular classification system is Bloom's Taxonomy.

Bloom's Taxonomy

Benjamin Bloom headed a group of educational psychologists who created a taxonomy for classifying objectives that commonly occur in educational settings. The taxonomy provides a useful structure in which to categorize test questions. The following is a brief description of sample verbs for thinking skills.

Sample Verbs for Thinking Skills:

- **Knowledge:** remember or recall information in approximately the same form in which it was learned; knowledge of terminology, specific facts (e.g., dates, events, places), or major ideas; mastery of subject matter

  Sample Verbs: collect, define, describe, identify, label, list, name, quote, show, tabulate, tell, when, where, who, etc.

- **Comprehension:** understand information based on prior learning; grasp meaning; translate knowledge into new context; interpret facts, compare, contrast; order, group, or infer causes; predict consequences

  Sample Verbs: convert, describe, differentiate, discuss, estimate, explain, extend, interpret, predict, restate, summarize, traces

- **Application:** use previously learned information in new and concrete situations to complete a problem or task; use methods, concepts, theories in new situations; solve problems using required skills or knowledge

  Sample Verbs: apply, complete, compute, demonstrate, discover, examine, illustrate, modify, relate, show, solve

- **Analysis:** break down informational materials into component parts; identification of components; organization of parts; seeing patterns; recognition of hidden meanings

Sample Verbs: analyze, categorize, compare, contrast, determine, differentiate, divide, examine, infer, separate

- **Synthesis:** apply prior knowledge and skills to produce a new or original whole; use old ideas to create new ones; generalize from given facts; relate knowledge from several areas; predict or draw conclusions

  Sample Verbs: combine, compose, create, design, devise, formulate, generalize, hypothesize, invent, integrate, modify, plan, rearrange, rewrite, what if?

- **Evaluation:** judge the value of material or methods for a given purpose; compare and discriminate between ideas; assess value of theories or presentations; make choices based on reasoned argument; verify value of evidence; recognize subjectivity

  Sample Verbs: assess, compare and contrast, conclude, convince, critique, decide, discriminate, explain, judge, justify, rank, recommend, select, summarize, support

Our goal is to balance items that call for lower level thinking skills (e.g., knowledge and comprehension) with items that call for higher level thinking skills (e.g., application, analysis, and synthesis). Since an important objective of K-12 instruction is to help students acquire and use higher level thinking skills, our items must assess the higher skills as well as the lower. Unless requested, items submitted should reflect a range of thinking skills. Item writers who submit items that utilize the higher order thinking skills are most useful in the test development process.

Avoid creating any items that could be labeled "so-what" questions. This kind of item, which is usually at the lowest level of thinking, asks for knowledge that has little or no value for assessing a student's mastery of a particular content standard. For example, in a reading comprehension test, items are not written that ask the student to recognize facts or details that are not relevant to a larger understanding of the text. An item that asks the student to identify the color of a person's jacket, for instance, would not be acceptable unless this fact is related to other ideas in the text. In other content areas, such as science and social studies, items are avoided that call for simple recall of unimportant facts. Occasionally, however, a particular content standard may require students to demonstrate recall of very specific information.

Avoid using mathematics items that call for computation without a relevant context or purpose related to the standard. For example, if a content standard requires students to understand the concept of perimeter, any required computation of perimeter should be simple rather than complicated, so that the item is testing the student's understanding of how to compute knowledge of perimeter rather than complex computation.

Many times, a low-level item can be rewritten so that it calls for higher levels of thinking. The following illustration of this process may help writers create items across the range of the taxonomy.

Acceptable low-level item:
Which of the following colors from the visible spectrum has wavelengths with the highest frequency?
A   red
B   violet
C   orange
D   blue

Acceptable higher-level item:
When visible light is passed through a glass prism, which color is refracted the most?
A   red
B   violet
C   orange
D   blue

The level of thinking required to answer a particular test item is not the same as the difficulty of that item. An item that calls for a student to make an analysis of an easy text or a familiar concept may be easier than one that calls for a student to demonstrate comprehension of a difficult text or a challenging concept.

Selecting Previously Published Passages

- Care should be taken to select passages for which the selection is self-contained. Each passage should function as a whole, having a beginning, middle, and an end.

- Passages should be well-written and contain accurate facts. It is also best to avoid passages containing nonstandard English.

- Passages should not come from textbooks or works that are frequently taught or highly anthologized.

- At each grade level, passages should represent a range of topics. Passages should be of probable interest to students at the target grade.

- Passages must contain enough information or ideas to serve as the basis for 10 to 12 test questions.

- Select passages from the Internet ONLY if they were previously published elsewhere as well.

- Passages must be of grade-appropriate length, depending upon the specifications of each program. The following general guidelines may be helpful:

K-2	100 to 400 words
3-4	200 to 600 words
5-6	200 to 800 words
7-8	200 to 900 words
9-12	300 to 1200 words

- Passages should be drawn from a variety of sources, including both magazines and books, unless the test specifications require otherwise. Do NOT locate passages exclusively from popular children's magazines.

- Passages should be selected to represent a variety of ethnic and cultural groups and include students from a variety of socioeconomic backgrounds.

- Passages should not contain content that suggests that the test developers are delving into personal values.

- Passages should not seek to advocate a particular value that may not be appropriate to all students taking the test.

27

- It is crucial that passages be free from problems of bias or sensitivity. Content should show males and females, various ethnic and cultural groups, the elderly, and the disabled in a variety of positive roles. Passages should not contain words that might be demeaning to a particular group or references that might tend to stereotype.

Guidelines for Specific Kinds of Passages

Narratives and folktales must have

- a recognizable theme

- round characters with sufficient motivation

- literary value (imagery, developing characters, theme, and suitable style)

- suitable setting (safe and supervised for younger children)

- appropriate vocabulary

- sufficient descriptive details

- some conflict and action (conflict should be resolved at the end)

- an attention-getting beginning and a satisfying ending

Personal narratives must have

- a recognizable theme

- a character with sufficient motivation

- literary value (imagery, developing characters, theme, and suitable style)

- suitable setting (safe and supervised for younger children)

- appropriate vocabulary

- sufficient descriptive details

- some conflict (either external or internal) and action (conflict should be resolved at the end)

- an attention-getting beginning and a satisfying ending

Informational, science, biography, persuasive, and multicultural passages must have

- an appropriate topic that is clearly presented

- connected thoughts with specific support

- a strong sense of order (chronological, cause/effect, comparison/contrast, classification or division,) where appropriate

- appropriate vocabulary, focus, and accurate details

- for commissioned pieces: two supporting documents (no more than one from the world wide web) with information used underlined and coded to text

- interesting and compelling information

- information focused on specific concepts, not scattershot

"How to" articles must have

- detailed and specific information

- appropriate sequential order

- a list of materials, where appropriate

- appropriate vocabulary and concepts for the grade level

- excellent organization

Bias and Sensitivity Issues

Educational Testing Service is committed to producing tests that provide equal opportunities for all students to demonstrate their knowledge and skills. Item writers must be keenly aware of potential issues of bias or sensitivity. Several important issues have been identified, as discussed below:

Stereotypical Representation

Items must avoid reinforcing stereotypes about specific groups. Examples include:

- Women should not always be shown in domestic situations, nor should they be portrayed as emotional, fearful, or incapable of making decisions.

- Items should not represent the assumption that minority populations are poor and uneducated.

- Native Americans should not be portrayed as more closely attuned to the natural world.

- Elderly individuals should not be represented as weak, hard-of-hearing, or muddled.

- Specific occupations should not be associated with certain groups (e.g., Chinese laundry).

Inclusion of Groups

Inclusion of the many cultural, ethnic, religious, and other groups in our society is extremely important in standards-based testing. Students should be able to "see themselves" in a test. Representation of groups can take several forms, including:

- illustrations that clearly show representations of genders and ethnic groups

- passages and other stimuli that illustrate and discuss multicultural topics

- use of passages and other stimuli whose authors are known to be or are identified as members of specific groups

- names in items that tend to be identified with specific ethnic groups

Bias Against Groups

An item may be biased if it contains content or language that is not familiar to specific groups of students. Examples include:

- geographical bias: Items cannot assume, for example, that students will know that it is customary to wear a scarf in cold weather.

- socioeconomic bias: Items should not expect that students can afford tickets to professional ball games or possess a computer.

- religious bias: Items should not assume that all students will have been to a Christian church or attend religious services.

- gender bias: Items should recognize that not all students are familiar with activities or concepts that have traditionally been associated with male or female members of society, including, for example, sports, cooking, clothing types, construction tools.

Use of Language

Some of the most subtle forms of bias occur in the use of language in testing. It is important to avoid terms that represent stereotypes and other demeaning assumptions about specific groups. Examples include:

- use of terms like lower class, housewife, red man

- use of occupational terms ending in man

- use of *he* or *mankind* to designate both genders

In addition, care must be taken that references to women or men or members of ethnic groups indicate the same level of respect in use of first names and titles.

Another form of bias in language use is sometimes termed "linguistic bias." Items that use idiomatic expressions or figurative language unnecessarily may be unfair to students whose first language is not English.

Topics or Concepts to Avoid

- Violence (including violence toward animals)
- Dying, death, disease
- Natural disasters with loss of life
- Drugs (illegal, prescription, alcohol, tobacco)
- Junk food
- Abuse, poverty, running away
- Divorce
- Socio-economic advantages (e.g., swimming pools, home computers, expensive vacations and purchases, etc.)
- Religion (including religious holidays: Halloween, Christmas, etc. Also do not write about birthdays)
- Complex discussions about sports
- Rap music, rock concerts
- Extra sensory perception, witchcraft, the supernatural
- Anything disrespectful, demeaning, moralistic, or chauvinistic
- Children coping with adult situations or decisions
- Unsafe practices or situations
- Losing a job, being fired
- Rats, roaches, lice
- Slavery
- Pornography
- War
- Evolution
- Dinosaurs and prehistoric times

Writers should avoid any topic that is likely to upset, or unduly distract, students and affect their performance on the test.

Glossary of Terms

Constructed-Response Items — Test items that require the student to create an individual response rather than to select from given responses. Constructed-response items may also be called short-answer items, short-response items, extended-response items, open-ended items, performance items, or performance tasks.

Distracters — Incorrect answers to a multiple-choice item.

Domain — The overall classification for the content standards within a subject area.

Item — A test question written in one of several possible item formats.

Item Format — The basic design of an item (e.g., multiple-choice, constructed-response).

Key — The correct answer.

Multiple-Choice Item — A stem plus (usually) four response options or answer choices.

Response Options — The (usually) four choices in a multiple-choice item, consisting of one answer and three distracters.

Rubric — The guidelines that indicate the range of responses and the score points used to score constructed-response items.

Standard — The statement that details what a student is expected to learn.

Stem — The initial part of the item in which the task is given. The stem may be a question, an incomplete statement, or a set of directions.

Stimulus — A passage, picture, graph, map, chart, quotation, or other text that students are asked to interpret when answering a test item.

Submission Checklist

When you are ready to submit your materials, please be sure you have completed the following:

- ❏ Read and understand the project's guiding documents (e.g., test and item specifications).

- ❏ Follow the guidelines in this Item Writer's Guide.

- ❏ Create original, interesting, grade-appropriate items that measure the standards.

- ❏ Use the review checklists for each item you are submitting to ensure it is of the highest quality.

- ❏ Complete the template provided for each item and follow submission guidelines.

- ❏ Submit source materials to verify factual information (if applicable).

- ❏ Follow appropriate security measures.

- ❏ Send items to ETS in a Word document via e-mail.

Helpful Resources

Graded Word Lists

Harris, A. J. & Jacobson, M. D. (1988). *Basic reading vocabularies.* New York: Macmillan.

Mogilner, A. (1992). *Children's writer's word book.* Cincinnati, OH: Writer's Digest Books.

Taylor, S. E., Frackenpohl, H., White, C., Nieroroda, B., Browning, C., & Birsner, E. P. (1989). *EDL core vocabularies in reading, mathematics, science, and social studies.* Austin, TX: Steck-Vaughn Company.

Bias and Sensitivity Information

Bias issues in test development. (1991). Amherst, MA: National Evaluation Systems, Inc.

Maggio, R. (1991). *The bias-free word finder: A dictionary of nondiscriminatory language.* Boston: Beacon Press.

Testing

ETS standards for quality and fairness. (2000). Princeton, NJ: Educational Testing Service.

Standards for educational and psychological testing. (1999). Washington, DC: American Educational Research Association.

Wilde, S. (2002). *Testing and standards: A brief encyclopedia.* Portsmouth, NH: Heinemann.

Thinking Skills

Bloom, B. S. (Ed.) (1956). *Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain.* New York; Toronto: Longmans, Green.

Marzano, R. J., Brandt, R. S., Hughes, C. S., Jones, B. F., Presseisen, B. Z., Rankin, S. C., & Suhor, C. (1988). *Dimensions of thinking: A framework for curriculum and instruction.* Alexandria, VA: Association for Supervision and Curriculum Development.

Style Concerns

University of Chicago Press. (1993). *The Chicago manual of style* (14th ed.). Chicago: Author.

Notes

REQUEST PERMISSION

Date:_____   Passage Code:_____

For Project:_____   Content Area:_____   Grade:_____

    Type of piece: Passage:_____   Excerpt:_____   Art:_____   Graph:_____   Other:_____

Passage Title: _____

By: _____

SECTION I –Where was piece found?

Title of book/source? _____

By (author/editors): _____

Published by:_____

Copyright owner/year: _____

SECTION II
If piece was found in an anthology, or permission was granted to the source where you found the passage,
PLEASE look at acknowledgments/Credits page and fill out the information below:

Book/Source where copyright is held: _____

Published by:_____

Copyright owner of passage/year: _____

Request permission for the following: Print:_____   Web:_____

Are edits needed: **YES / NO** (be sure to attach copy of piece with requested edits)

Does passage include art? **YES / NO**   Do you want permission to use art? **YES / NO**

Other Needs: _____

Please make sure all copyright pages are included.

Piece_____   Cover Page_____   Copyright page_____   Acknowledgement's pages if applicable_____

Requester's Signature:_____ Content Coordinator: _____

Request for Permission

Date:

TO:

Introductory paragraph describing program

We are writing to ask for non-exclusive U.S. permission to reprint the following for use on this test:

Title: «PassageExcerpt_Title»

By: «Author»

From: «book_title_passage_is_originally_from»

Copyright year: «Copyright_year_of_passage»

If originally found elsewhere: «Passage_Found_In»

We need permission for:

Print: _____ Web: ____

ART_____ to re-illustrate:_____

A photocopy of the requested materials is attached, with any requested edits indicated. If any edits have been requested, it is because we need to condense the material for the students to take the test in the allotted time, or because specific words or topics may not be acceptable within the assessment testing industry. If edits or excerpts are not allowed, please let us know as soon as possible.

Publication Dates: The selection(s) will first be placed in a field test edition in 20XX. If the results are satisfactory, the selection will then be used in an operational test after the satisfactory results and before December 20XX.

Length of Non-Exclusive U.S. Licensure:

Total Estimated Print Run During Licensure:

Field: «FT_Print_Run»

Operational: «OP_Print_Run»

Total Print Run: «Total_Print_Run»

Approximate number of pages in a soft-cover test booklet:

Sale Price $

If you require payment for this permission, please fill out the enclosed W-9 form. (If a W-9 form is not provided it is because we already have your company in our system.) Unless you request otherwise, we will provide any payment upon the date of publication. We will acknowledge the permission request in a conventional form, indicating the copyright holder and including author, title, publisher, and year of publication.

If you do not control the rights to this material, or if we should direct this request to another party, please let us know as soon as possible.

We thank you for your consideration. We ask that, if possible, you use the form provided on the following page(s) and forward your permission to the above address or fax it to **XXX-XXX-XXXX**. Please contact me, with any questions you may have. My direct telephone number is **XXX-XXX-XXXX**, and my email address is XXXXXX@XXX.

Sincerely,

Permissions Editor

Enclosure:

Sample Content Review Form

Name _____   Date _____   Form Number _____

Directions: For the following questions, leave blank if the answer is "yes" or put a checkmark to indicate a concern.

Item Number	Correct Answer?	Correct Standard?	Match to Item Specs?	Appropriate Difficulty Level Indicated?	Reflects Classroom Instruction?	Free from Bias or Sensitivity Issues?	Item Status • Use as is • Revise and edit • Do not use
1							
2							
3							
4							
5							
6							
7							
8							

Sample Bias and Sensitivity Review Form

Name _____   Date _____   Form Number _____

Item Number	No Bias or Sensitivity Issues	Contains Bias or Sensitivity Issues (Note type of bias and provide comments.) • Gender • Racial • Ethnic • Religious • Age • Socio-Economic Status • Cultural • Disability • Inappropriate Behaviors • Other	Suggestions for Revision
1			
2			
3			
4			
5			

Sample Data Review Form

Name _____   Date _____   Form Number _____

Directions: For the following item data categories, leave blank if the data are within the appropriate limits or put a checkmark to indicate a concern. Put an X to indicate the status of each item.

Item Number	Item Data				Item Status			
	P-Value	Response Distribution	Point-Biserial	Mantel-Haenszel	Use As Is	Revise and Use	Revise and Re-Field Test	Reject
1								
2								
3								
4								
5								
6								
7								
8								

Item Writer:		Contract #:		Project #:	
Content Area:					
Grade Level:		Correct Answer:		Thinking Skill:	
Standard Code:					
Domain:					
Category:					
Standard:					
Item Code:					

Type item here.

A   xxx
B   xxx
C   xxx
D   xxx

6 Which (is, are) never present when lempens are being vexed?

    A Ram and krem
    B Krem and pommel
    C Tevar and krem
    D Krem

Reason: _____

7 The seaming function of the portar is most effectively used in connection with

    A a nova brig.
    B a raver.
    C the notrin.
    D the nimmer tag.

Reason: _____

8 The total profit for Grezzle, Inc. is less than

    A 10 million.
    B 8 million.
    C 5 million.
    D 1 million.

Reason: _____

9 The pavin will typically estivate in which of these situations?

    A When the jab strikes the vap
    B When the voble mavils
    C When the betam is whant
    D When the rapple torks, if the rapple is galled or twizzled

Reason: _____

10 When exposed to yambs, the limpa will

    A
    B
    C
    D

Reason: _____

Putting Your Test Items to the Test

	Right Reason	Wrong Reason
Correct Student Response		
Incorrect Student Response		